ΚΕΦΑΛΑΙΟ 2ο
ΕΙΣΑΓΩΓΙΚΑ ΠΕΡΙ ΣΤΑΤΙΣΤΙΚΗΣ

Πανεπιστήμιο Αιγαίου, Ακαδημαϊκό Έτος 2023-2024

Ορισμός Στατιστικής

Ορισμός: (Βόντα & Καραγρηγορίου 2012) **Στατιστική** είναι ο επιστημονικός κλάδος όπου διατυπώνονται τα αξιώματα και η μεθοδολογία που διέπουν

(α) το σχεδιασμό και τον τρόπο συλλογής δεδομένων και πληροφοριών

(β) την οργάνωση, κατανομή και ανάλυση δεδομένων και

(γ) τη διατύπωση συμπερασμάτων, αποτελεσμάτων και αποφάσεων.

Οι **κλάδοι της στατιστικής** που καταπιάνονται με τα παραπάνω 3 θέματα είναι:

(α) Σχεδιασμός Πειραμάτων (Experimental Design)

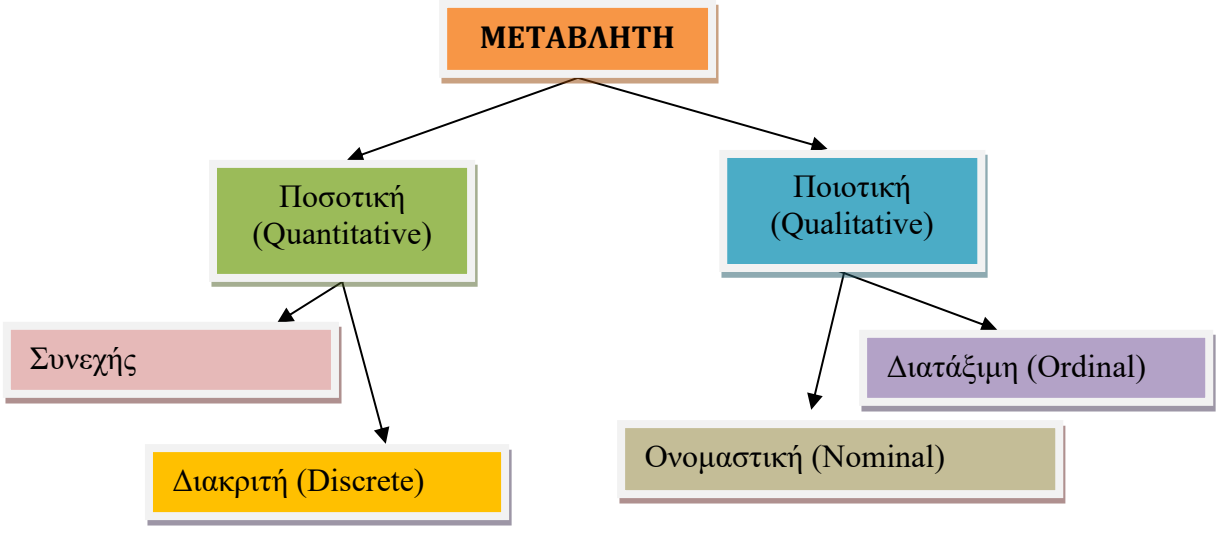
(β) Περιγραφική Στατιστική (Descriptive Statistics)

(γ) Στατιστική Συμπερασματολογία (Inferential Statistics)

Αντικείμενο της Στατιστικής είναι η εξαγωγή συμπερασμάτων για έναν πληθυσμό βάση των στοιχείων που μπορούμε να πάρουμε από ένα δείγμα. Η Στατιστική ασχολείται με το σχεδιασμό του τρόπου συλλογής των δεδομένων, έτσι ώστε τα συμπεράσματα που εξάγονται να είναι το δυνατό πιο ακριβή.

Επίσης, ενδιαφέρει η μελέτη κάποιου συγκεκριμένου χαρακτηριστικού ενός πληθυσμού (π.χ. ηλικία, βάρος, επίπεδο μόρφωσης, ποσοστό ψηφοφόρων, ποσοστό ελαττωματικών εξαρτημάτων, αριθμός ασθενών κλπ). Συνήθως δεν είναι εφικτή η εξέταση όλων των μελών ενός πληθυσμού και γι'αυτο το λόγο, η μελέτη επικεντρώνεται στο **δείγμα**, το οποίο αποτελεί ένα μικρό (σχετικά) τμήμα του πληθυσμού. Όταν γίνεται εξέταση όλων των μελών ενός πληθυσμού, τότε κάνουμε απογραφή του πληθυσμού.

Με κάθε χαρακτηριστικό υπό μελέτη, συνδέεται και μια μεταβλητή (variable) η οποία λαμβάνει διάφορες τιμές (και άρα μεταβάλλεται) ανάλογα με τον "ερωτώμενο".



Παραδείγματα:

Ονομαστικές (Nominal) μεταβλητές: Λαμβάνουν μη αριθμητικές τιμές, οι οποίες δεν επιδέχονται διάταξης (π.χ., θρήσκευμα, φύλο, χρώμα).

Κατηγορικές (Ordinal) μεταβλητές: Λαμβάνουν μη αριθμητικές τιμές, οι οποίες επιδέχονται διάταξης (π.χ. Εισοδηματική τάξη με τιμές στο {"Κατώτατη", "Κατώτερη", "Μεσαία", "Ανώτερη", "Ανώτατη"}, Στάδια ασθένειας, Γνώμη με τιμές στο {"Πολύ Κακή", "Κακή", "Μέτρια", "Πολύ Καλή", "Άριστη"}.

Διακριτές (Discrete) μεταβλητές: Λαμβάνουν αριθμήσιμο πλήθος αριθμητικών τιμών (π.χ. αριθμός παιδιών σε μια οικογένεια, αριθμός ασθενών ανά ημέρα, αριθμός τηλεφωνημάτων ανά ώρα, αριθμός άστοχων ελεύθερων βολών, αριθμός ελαττωματικών αντικειμένων κλπ).

Συνεχείς (Continuous) μεταβλητές: Λαμβάνουν υπερ-αριθμήσιμο πλήθος αριθμητικών τιμών (συνήθως, τιμές σε ένα διάστημα π.χ. ταχύτητα, θερμοκρασία, ύψος, βάρος, δείκτης μάζας/σώματος, αρτηριακή πίεση, επιτόκιο καταθέσεων, τιμή κλεισίματος μετοχής, τιμή κλεισίματος γενικού δείκτη χρηματιστηρίου κλπ).

- Τις μεταβλητές τις συμβολίζουμε με κεφαλαία γράμματα (π.χ. X, Y, Z κλπ ή X_1, X_2,..., Y_1, Y_2, ... κλπ)
- Οι συγκεκριμένες αριθμητικές εκφράσεις των μεταβλητών ονομάζονται τιμές των μεταβλητών και συμβολίζονται με μικρά γράμματα (π.χ., x, y, z κλπ ή x_1, x_2, ..., y_1, y_2, ... κλπ).

Παράδειγμα:

i	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
x_i	55	59	47	44	52	60	62	57	56	66

- Από τον παραπάνω πίνακα, βλέπουμε ότι έχουμε ένα δείγμα 10 τιμών, όπου για το 1ο μέλος του δείγματος, η τιμή του χαρακτηριστικού (μεταβλητή) *X* είναι 55, δηλ. $x_1 = 55$. Όμοια, για το 2ο μέλος του δείγματος, η τιμή του χαρακτηριστικού *X* είναι 59, δηλ. $x_2 = 59$ κ.ο.κ.
- Σημαντικό ρόλο παίζει και το *διατεταγμένο* δείγμα, δηλ. το δείγμα στο οποίο οι τιμές του αρχικού δείγματος έχουν διαταχθεί από τη μικρότερη προς τη μεγαλύτερη.
- Συμβολισμός: Ως $x_{(i)}$ συμβολίζουμε την *i*-οστη διατεταγμένη παρατήρηση
- $x_{(1)}$: Είναι η ελάχιστη παρατήρηση, $x_{(n)}$: Είναι η μέγιστη παρατήρηση

i	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
$x_{(i)}$	44	47	52	55	56	57	59	60	62	66

Περί Δειγματοληψίας

- Η αμεροληψία είναι ίσως το πιο σημαντικό στοιχείο που πρέπει να ληφθεί υπόψη κατά την επιλογή των μελών του πληθυσμού, τα οποία θα αποτελέσουν το δείγμα.
- Μεροληπτική δειγματοληψία οδηγεί σε εσφαλμένα συμπεράσματα και απώλεια αξιοπιστίας.
- **Αποφυγή μεροληψίας;** Επιλογή Απλού Τυχαίου Δείγματος (*Simple Random Sample*).
- **Πως;** Τοποθετούνται όλα τα μέλη του πληθυσμού σε μια κάλπη, ανακατεύονται καλά (ώστε να διασφαλιστεί ότι κάθε μέλος έχει την ίδια πιθανότητα επιλογής) και στη συνέχεια επιλέγονται τα μέλη του πληθυσμού που θα απαρτίζουν το δείγμα.

Κυριότερες δειγματοληπτικές τεχνικές

- Απλή Τυχαία Δειγματοληψία (*simple random sampling*)
- Στρωματοποιημένη Δειγματοληψία (*stratified sampling*)
- Δειγματοληψία κατά ομάδες (*cluster sampling*)
- Συστηματική δειγματοληψία (*systematic sampling*)

- Για περισσότερες λεπτομέρειες, δείτε Βόντα & Καραγρηγορίου (2012)[1].
- Στο μάθημα αυτό, δε θα μας απασχολήσει περαιτέρω ο τρόπος επιλογής ενός τυχαίου δείγματος.

[1] *Εφαρμοσμένη Στατιστική Ανάλυση & Στοιχεία Πιθανοτήτων*. Συγγραφείς: Ι. Βόντα, Α. Καραγρηγορίου, εκδότης ΜΑΡΙΝΗΣ ΣΠΥΡΟΣ & ΣΙΑ Ο. Ε.

Τρόποι Συλλογής Δεδομένων

- Ελληνική Στατιστική Αρχή (ΕΛ. ΣΤΑΤ., www.statistics.gr)
- Ευρωπαϊκή Στατιστική Αρχή (Eurostat, http://ec.europa.eu/eurostat/web/main/home)
- Απευθείας από Δημόσιες Υπηρεσίες, Ο.Τ.Α.
- Από ιδιωτικές εταιρείες.
- Μέσω διαδικτύου, π.χ.
  - Word bank Open Data (https://data.worldbank.org/)
  - Yahoo!finance (https://finance.yahoo.com)
  - World Health Organization (https://www.who.int/en/)
  - UCI Machine Learning Repository (https://archive.ics.uci.edu/ml/index.php)
- **Δεν ξεχνάμε να αναφέρουμε την πηγή από την οποία πήραμε τα δεδομένα!!**